Consistency Constraints for Overlapping Data Clustering

Jakob Hansen
Applied Mathematics and Computational Science
University of Pennsylvania
209 S. 33rd Street, Philadelphia, PA 19104
jhansen@sas.upenn.edu

Jared Culbertson
Sensors Directorate
Air Force Research Laboratory
2241 Avionics Circle, WPAFB, OH 45433
jared.culbertson@us.af.mil

Peter F. Stiller
Department of Mathematics
Texas A&M University
MS3368, College Station, Texas 77843
stiller@math.tamu.edu

Dan P. Guralnik
School of Engineering and Applied Science
University of Pennsylvania
200 S. 33rd St., Philadelphia, Pennsylvania 19104
guraldan@seas.upenn.edu

***Abstract*—We examine overlapping clustering schemes with functorial constraints, in the spirit of Carlsson–Mémoli. This avoids issues arising from the chaining required by partition-based methods. Our principal result shows that any clustering functor is naturally constrained to refine single-linkage clusters and be refined by maximal-linkage clusters. We work in the context of metric spaces with non-expansive maps, which is appropriate for modeling data processing which does not increase information content. Examples of a number of such clustering algorithms are given in the text.**

***Index Terms*—functorial clustering, clustering with overlaps, metric clustering**

I. Introduction

The problem of data clustering has been extensively studied. Clustering is used in fields as diverse as biology, psychology, machine learning, sociology, image processing, and chemistry, in order to discover hidden structure in data. Among the earliest systematic treatment of clustering theory was that of Jardine and Sibson in 1971 [21]. Since then, there have been several distinct directions of research in clustering theory, with only modest communication between researchers pursuing different paths.

The classical work of Jardine and Sibson was followed by other similarly comprehensive works such as Everitt [17]. Further theoretical work on these mostly classical methods was also done by Kleinberg [22] and Carlsson and Mémoli [9], [10]. Work on computing phylogenetic trees inspired a seminal paper by Bandelt and Dress [3] on split decompositions of metrics. This line of research was continued with investigations into split systems and cut points of injective envelopes of metric spaces. Representative papers include [15] and [16]. While not explicitly clustering methods, these methods are quite similar in spirit to stratified clustering schemes. In this category we might also add the classification of injective envelopes of six-point metric spaces by Sturmfels and Yu [26].

The authors are grateful for the financial support of the Air Force Office of Scientific Research under the LRIR 12RY02COR, LRIR 15RYCOR153, and MURI FA9550-10-1-0567 grants.

Bandelt and Dress also had a large influence on another field as a result of their work on weak hierarchies [2], [4]. This led to work by Diatta, Bertrand, Barthélemy, Brucker, and others on indexed set systems (see, e.g., [6], [8], [14]). An interesting recent development here is the work by Janowitz on ordinal clustering [19].

Additional work has been done on topologically-based clustering methods. This includes the `Mapper` algorithm by Singh, Mémoli, and Carlsson [25], as well as work on persistence based methods [12] and Reeb graphs [18].

Meanwhile, most users of clustering methods default either to a classical linkage-based clustering method (such as single-linkage or complete-linkage) or to more geometrically based methods like k-means. Unfortunately, the wide array of clustering theories has had little impact on the actual practice of clustering.

This paper works to bridge some of these gaps by extending a recent paper of Carlsson and Mémoli [10]. Their paper introduced the idea of viewing clustering methods as functors from a category of metric spaces to a category of classifying objects giving rise to clusters (e.g. partitions, dendrograms). We will only make use of the very basics of category theory including the notions of categories, morphisms, functors and natural transformations. This abstract language is extremely powerful for not only compactly representing complex information, but also providing a formalism for reasoning about natural operations. For those unfamiliar with these concepts, see either [23] (for a mathematical treatment) or [5] (for a computer science perspective). Many desirable properties of a clustering method are subsumed in functoriality when morphisms are properly chosen. One of our principal goals is to extend their theory of functorial clustering schemes to methods that allow overlapping clusters, and in so doing obviate some of the unpleasant effects of chaining that occur for example with single-linkage. Rather than relying on chaining to overcome certain technical problems, we accept overlapping clusters.

ISBN: 1-60132-502-9, CSREA Press ©

II. Definitions

Let X be a set, to be thought of as a set of unlabeled data to be analyzed. In order to make as few assumptions as possible, we only require that X be endowed with a metric, which we will often refer to as d_X. However, we do not assume, for example, that X is embeddable in Euclidean space or that X is obtained by sampling from some distribution. Recall that a *cover* $\mathcal{C}$ of X is a collection of subsets of X whose union is X. A cover $\mathcal{C}$ of X is a *partition* of X if $A \cap B = \varnothing$ for any pair of distinct subsets $A, B \in \mathcal{C}$. Also recall that a cover $\mathcal{C}_1$ is said to *refine* a cover $\mathcal{C}_2$, if every $U \in \mathcal{C}_1$ is contained in some $V \in \mathcal{C}_2$.

Traditionally, a clustering method applied to an input dataset is expected to produce a partition of X. The work by Kleinberg [22] highlights the need for a rigorous treatment of the formal relations between the non-expansive maps among finite metric spaces on the one hand, and refinement relations among partitions produced by distance-based clustering methods on the other. In fact, the main result of *loc. cit.* clearly states refinement relations as the obstruction to the "richness" axiom (stating that every partition be obtainable as an output of the clustering method for some suitably chosen input, this axiom seems to us as the single least debatable of Kleinberg's axioms).

Accepting the philosophical position of Carlsson and Mémoli [9] that functoriality of the clustering map is a suitable replacement for the rest of Kleinberg's axioms, our interest in clustering *with overlaps* leads us to formulating a restriction on the class of covers of X acceptable as outputs of a distance-based clustering method. However, we prefer to view functoriality as a way of imposing constraints on consistent clustering across datasets, rather than as a set of axioms that must be adhered to.

Following Jardine and Sibson [21], we consider a clustering as encoded by a symmetric and reflexive relation R, with clusters being defined as the fibers, $[x]_R := \{y \,|\, xRy\}$ of the relation. This point of view shows that, in addition to the functoriality constraints already mentioned, a clustering method affording overlaps requires a weakening of the transitivity property (characteristic of partitioning methods). Should transitivity be dropped completely, all that remains is the observation that the fibers of R form a cover of X. Still, intuitively, for the purpose of distance-based clustering one feels that three points x, y, and z, which are pairwise "similar" to some (measurable) degree need to be regarded as "jointly" similar to the same degree. Likewise, this observation should remain valid for larger set of points. This motivates the following definition:

Definition 1. Let X be a non-empty finite set. A *non-nested flag cover* (or simply *flag cover*) is a cover $\mathcal{C}$ of X satisfying the following conditions:

(1) If $A, B \in \mathcal{C}$ and $A \subseteq B$, then $A = B$.
(2) The abstract simplicial complex $\mathbb{K}(\mathcal{C})$ consisting of all the subsets of elements of $\mathcal{C}$ is a flag complex.

We denote the set of flag covers of X by $\mathbf{Cov}_{\text{Þ}}(X)$. □

Note that $\mathcal{C}$ is the collection of maximal simplices of $\mathbb{K}(\mathcal{C})$, with a simplex spanning $S \subseteq X$ if and only if S is contained in some element of the cover $\mathcal{C}$. Thus, $\mathbb{K}(\mathcal{C})$ is flag if, for every $S \subseteq X$, S spans a simplex in $\mathbb{K}(\mathcal{C})$ whenever every pair of distinct points $x, y \in S$ is contained in an element of $\mathcal{C}$. In particular, every partition of X is a flag cover of X.

Finally, note that any cover $\mathcal{U}$ of X can be "upgraded" to a non-nested flag cover $\mathcal{U}^*$, commonly referred to as the *flagification of* $\mathcal{U}$, in a minimal way, where $\mathcal{U}$ will refine $\mathcal{U}^*$ and $\mathcal{U}^*$ will refine any other flag cover which $\mathcal{U}$ refines. This may be done by iteratively adjoining to $\mathcal{U}$ any clusters mandated by the flag condition, and then removing all the non-maximal ones.

Perhaps the most common notion in the clustering literature (see, e.g., [21], [14]) related to flag complexes is that of *maximally linked sets*. We recall:

Definition 2. Let X be a set and let R be a symmetric, reflexive relation $R \subset X \times X$. A subset $S \subseteq X$ is *maximally linked with respect to* R if (1) xRy for all $x, y \in S$, and (2) S is not properly contained in any subset of X satisfying (1). □

Clearly, picking $\mathcal{C}$ to be the set of all maximally linked subsets of X with respect to R results in a flag cover of X. One of the most studied constructions of this form is the *Vietoris–Rips complex*, arising from a metric space (X, d_X) as $\mathbb{K}(\mathcal{C})$ upon setting $x_1 R x_2 \Leftrightarrow d_X(x_1, x_2) \leq \delta$, for some $\delta \geq 0$.

Definition 3. A *persistent cover* on X is a function $\theta_X \colon \mathbb{R}_{\geq 0} \to \mathbf{Cov}_{\text{Þ}}(X)$ such that

(1) If $t_1 \leq t_2$ then $\theta_X(t_1)$ refines $\theta_X(t_2)$.
(2) For any t, there is an $\varepsilon > 0$ with $\theta_X(t') = \theta_X(t)$ for all $t' \in [t, t + \varepsilon)$.

If we also have (3) below, then we call θ_X a *sieve* on X:

(3) There exists $t \in \mathbb{R}_{\geq 0}$ such that $\theta_X(t)$ is the trivial cover $\{X\}$. □

Persistent covers and sieves are a direct generalization of Carlsson and Mémoli's *persistent sets* and *dendograms*, which satisfy the same conditions, but have the set of partitions of X as codomain. They may also be seen as a sort of strictly isotone indexed set system as in [7], where the index of each set $A \in \bigcup_{t \in \mathbb{R}_{\geq 0}} \theta_X(t)$ is given by the infimum of the values of t such that $A \in \theta_X(t)$.

We now consider the category $\mathbf{Met}$, which has finite metric spaces as objects and non-expansive mappings as morphisms. That is, a map of sets $f \colon X \to Y$ is a morphism $(X, d_X) \to (Y, d_Y)$ in $\mathbf{Met}$ if for any $x, x' \in X$, $d_Y(f(x), f(x')) \leq d_X(x, x')$. This is the same as saying $f^*(d_Y) \leq d_X$, where $f^*(d_Y)$ is the metric on X given by $f^*(d_Y)(x_1, x_2) = d_Y(f(x_1), f(x_2))$. Note that any morphism $f \colon (X, d_X) \to (Y, d_Y)$ factors through $(X, f^*(d_Y))$. We abuse terminology somewhat by allowing zero distances between points in our finite metric spaces. In this way $(X, f^*(d_Y))$ is a valid object in $\mathbf{Met}$ even when $f \colon X \to Y$ is not injective.

We want to take objects in $\mathbf{Met}$ and convert them into (collections of) clusters in various ways. The category $\mathbf{Cov}_{\text{Þ}}$

is the category of ordered pairs $(X, \mathcal{C})$, where X is a set and $\mathcal{C}$ is a flag cover of X. A morphism between $(X, \mathcal{C})$ and $(Y, \mathcal{D})$ is a map of sets $f\colon X \to Y$ such that $\mathcal{C}$ is a refinement of $f^{-1}(\mathcal{D})$. These are called *consistent maps*. Note that $f^{-1}(\mathcal{D})$ need not be a flag cover of X, though it becomes one upon removal of its non-maximal elements.

The category $\mathbf{Part}$ of partitions is a subcategory of $\mathbf{Cov}_{\triangleright}$, where only coverings that are also partitions are allowed. We define $\mathbf{Sieve}$ as the category of pairs (X, θ_X), where $\theta_X\colon \mathbb{R}_{\geq 0} \to \mathbf{Cov}_{\triangleright}(X)$ is a sieve on X. The morphisms in $\mathbf{Sieve}$ are an extension of the morphisms of $\mathbf{Cov}_{\triangleright}$; that is, a set map $f\colon X \to Y$ is a morphism of sieves $(X, \theta_X) \to (Y, \theta_Y)$ if for every $t \in \mathbb{R}_{\geq 0}$, $\theta_X(t)$ refines $f^{-1}(\theta_Y(t))$. Note that this means that we have a family of functors from $\mathbf{Sieve}$ to $\mathbf{Cov}_{\triangleright}$, by restricting to a particular value of the parameter t. For convenience, we summarize these categories in Figure 1.

Category	Objects	Morphisms
$\mathbf{Met}$	Finite metric spaces (X, d_X)	Non-expansive maps
$\mathbf{Met}^{\text{inj}}$	Finite metric spaces (X, d_X)	Non-expansive injections
$\mathbf{Part}$	$(X, \mathcal{P}_X)$, $\mathcal{P}_X$ a partition of X	Consistent maps
$\mathbf{Cov}_{\triangleright}$	$(X, \mathcal{C}_X)$, $\mathcal{C}_X$ a flag cover of X	Consistent maps
$\mathbf{PerSet}$	(X, Φ_X), Φ_X a persistent set on X	Consistent maps
$\mathbf{Dendro}$	(X, Φ_X), Φ_X a dendrogram on X	Consistent maps
$\mathbf{PerCov}$	(X, θ_X), θ_X a persistent cover of X	Consistent maps
$\mathbf{Sieve}$	(X, θ_X), θ_X a sieve on X	Consistent maps

Fig. 1. Summary description of relevant categories. A consistent map from $(X, \mathcal{C}_X)$ to $(Y, \mathcal{C}_Y)$ is a set function $f\colon X \to Y$ such that for every $A \in \mathcal{C}_X$, there exists some $B \in \mathcal{C}_Y$ such that $A \subseteq f^{-1}(B)$. A consistent map from (X, θ_X) to (Y, θ_Y) is a set function f such that for every $t \in \mathbb{R}_{\geq 0}$, f is a consistent map from $(X, \theta_X(t))$ to $(Y, \theta_Y(t))$.

III. Flat Clustering

The primary development of this paper focuses on clustering methods that work at a fixed scale, giving clusters of similar data points either as blocks of a partition or sets in a covering. In the next section, we will briefly describe how these methods can be extended to hierarchical versions.

A. Functors on Met

We consider a *flat* or *nonhierarchical (overlapping) clustering* to be a covariant functor $\mathbf{F}$ from $\mathbf{Met}$ to $\mathbf{Cov}_{\triangleright}$, which restricts to the identity on the underlying set, i.e., $\mathbf{F}(X, d_X)$ takes the form $(X, \mathcal{C})$ where $\mathcal{C}$ is a flag cover of X. We will refer to such $\mathbf{F}$ as *clustering functors*. A reasonable first question is whether there are any interesting such functors, and the following definition provides a useful way of constructing many examples.

Definition 4. Let $\mathcal{T}$ be a set of finite metric spaces. Given a metric space (X, d_X), define a relation R on X with xRy if and only if there exists a morphism t from some $T \in \mathcal{T}$ into (X, d_X) satisfying $x, y \in \mathrm{Im}(t)$. Let $\mathbf{ML}^{\mathcal{T}}(X, d_X)$ be the covering of X by maximally linked subsets under R. We refer to $\mathbf{ML}^{\mathcal{T}}$ as the *clustering functor generated by* $\mathcal{T}$.

Remark 5. Clearly, the relation R above is preserved under any morphism. By this we mean that if we have a morphism $f\colon X \to Y$ in $\mathbf{Met}$ and xRy, so that there is a morphism $t\colon T \to X$ for some $T \in \mathcal{T}$ with x and y in the image of t, then the composition $f \circ t\colon T \to Y$ yields $f(x)Rf(y)$. This verifies that $\mathbf{ML}^{\mathcal{T}}$ is, indeed, functorial for $\mathbf{Met}$.

A wide range of clustering functors are obtainable in this way. We begin with the standard single-linkage clustering scheme. Given a parameter δ, we can construct the Vietoris–Rips complex from any metric space X by adding an edge between two points whenever the distance between them is at most δ. Define $\mathcal{C}_{\mathbf{SL}}$ as the partition of X given by the connected components of the Vietoris–Rips complex of X, and define $\mathbf{SL}_\delta\colon \mathbf{Met} \to \mathbf{Cov}_{\triangleright}$ by $(X, d_X) \mapsto (X, \mathcal{C}_{\mathbf{SL}})$. Carlsson and Mémoli, in [10], have shown that $\mathbf{SL}_\delta$ is functorial when viewed as a map to $\mathbf{Part}$. Since $\mathbf{Cov}_{\triangleright}$ contains $\mathbf{Part}$, the mapping $\mathbf{SL}_\delta$ is also functorial with $\mathbf{Cov}_{\triangleright}$ as target category. Alternatively, it is easy to see that $\mathbf{SL}_\delta$ is generated by the collection of spaces $\Lambda^k_\delta = \{0, \ldots, k\}$ endowed with the metric $d(i, j) = \delta|i - j|$, with k ranging over the positive integers (see Definition 4).

We now define maximal-linkage clustering in a similar fashion. Given δ, again construct the Vietoris–Rips complex of X. We then take $\mathcal{C}_{\mathbf{ML}}$ to be the set of maximal simplices of this complex. We define $\mathbf{ML}_\delta$ as the map taking (X, d_X) to $(X, \mathcal{C}_{\mathbf{ML}})$. Alternatively, $\mathbf{ML}_\delta$ is $\mathbf{ML}^{\mathcal{T}}$ for $\mathcal{T} = \{\Lambda^1_\delta\}$.

Theorem 6. *$\mathbf{ML}_\delta$ is a surjective functor from $\mathbf{Met}$ to $\mathbf{Cov}_{\triangleright}$.*

Proof. The image under $\mathbf{ML}_\delta$ of a morphism f in $\mathbf{Met}$ should be the morphism in $\mathbf{Cov}_{\triangleright}$ given by the same set function, if it is indeed a morphism in $\mathbf{Cov}_{\triangleright}$. Thus, as long as $\mathbf{ML}_\delta$ maps morphisms to morphisms, it will respect composition. In the following diagram, we need to show that $\mathbf{ML}_\delta(f)$ is a morphism in $\mathbf{Cov}_{\triangleright}$ given that f is a morphism in $\mathbf{Met}$.

$$\begin{array}{ccc} (X, d_X) & \xrightarrow{\quad f \quad} & (Y, d_Y) \\ \Big\downarrow{\scriptstyle \mathbf{ML}_\delta} & & \Big\downarrow{\scriptstyle \mathbf{ML}_\delta} \\ (X, \mathcal{C}_X) & \xrightarrow[\quad \mathbf{ML}_\delta(f) \quad]{} & (Y, \mathcal{C}_Y) \end{array}$$

Define a reflexive symmetric relation D_X on X with xD_Xx' if $d_X(x, x') \leq \delta$; similarly, let yD_Yy' if $d_Y(y, y') \leq \delta$. Then for any morphism f in $\mathbf{Met}$, $xD_Xx' \Rightarrow f(x)D_Yf(x')$. Under a mild abuse of notation, this means that $D_X \subseteq f^{-1}(D_Y)$. Note that the sets in $\mathcal{C}_X$ are the maximal linked sets under D_X. Since $f^{-1}(D_Y)$ contains D_X, every maximal linked set under D_X is contained in a maximal linked set under $f^{-1}(D_Y)$. Hence $\mathcal{C}_X$ refines $f^{-1}(\mathcal{C}_Y)$.

An alternative proof of this fact is to note that the sets in $\mathcal{C}_X$ are the maximal linked sets, i.e. if U is one such set then $d_X(x_1, x_2) \leq \delta$ for every $x_1 \in U$ and $x_2 \in U$, and for every $\tilde{x} \in X, \tilde{x} \notin U$ there is an $x \in U$ with $d_X(x, \tilde{x}) > \delta$. It follows that $d_Y(f(x_1), f(x_2)) \leq \delta$ so that all the points in $f(U)$ are within δ of each other. They therefore lie in some maximal

linked set V in $\mathcal{C}_Y$. It follows that $U \subset f^{-1}(V)$ and that $\mathcal{C}_X$ refines $f^{-1}(\mathcal{C}_Y)$.

To see that $\mathbf{ML}_\delta$ is surjective, note that the cover $\mathcal{C}$ implicitly defines a simplicial complex on X by taking the sets in the cover as maximal simplices. Because $\mathcal{C}$ is a flag cover, this complex is flag, uniquely determined by its 1-skeleton. We can therefore metrize the 1-skeleton of this complex by setting every edge length to δ, and setting the distance between any two disconnected points to be 2δ. Then the distance between two points in the complex is less than or equal to δ if and only if they are in the same simplex. This implies that the maximal simplices are exactly the maximal linked sets under this metric. Thus every flag cover arises from some metric on X under the $\mathbf{ML}_\delta$ map, so that this map is surjective. □

The concept in the preceding proof of defining a symmetric reflexive relation on X and then taking its maximal linked sets is an important one. We may reformulate $\mathbf{SL}_\delta$ in terms of a relation by letting $x \sim_\delta y$ for two elements x, y of X if there is a positive integer k and a sequence of points $x = x_0, x_1, \ldots, x_{k-1}, x_k = y$ such that $d(x_{i-1}, x_i) \leq \delta$ for all $1 \leq i \leq k$. This is just a more explicit way of saying that there is a non-expansive map $\Lambda^k_\delta \to X$ containing x, y in its image (see the definition of $\mathbf{SL}_\delta$ above). In this case the relation is an equivalence relation and the single-linkage clusters are simply the equivalence classes of $\sim_\delta$. Similarly, $\mathbf{ML}_\delta$ consists of the maximal linked sets of the relation M where xMy if $d(x, y) \leq \delta$.

This suggests the possibility of expanding the relation M to include more pairs but not to the extent of the relation $\sim_\delta$. One potential method is to fix a positive integer k and define a relation R^k_δ as before, such that $xR^k_\delta y$ if there exists a sequence of $k+1$ points (not necessarily all distinct) $x = x_0, x_1, \ldots x_{k-1}, x_k = y$ such that $d(x_{i-1}, x_i) \leq \delta$ for $1 \leq i \leq k$. In other words, we can get from x to y in k steps of size at most δ. We denote the resultant map from $\mathbf{Met} \to \mathbf{Cov}_\blacktriangleright$, given by taking maximal linked sets of R^k_δ, as $\mathbf{L}^k_\delta$, and call it *k-linkage clustering*. Observe $\mathbf{L}^k_\delta$ is generated by $\mathcal{T} = \{\Lambda^k_\delta\}$. In particular, $\mathbf{L}^k_\delta$ is a functor by Remark 5.

Of course, additional relations are possible. For example, we could also define R^K_δ where we can take as many steps as we like provided that the sum of the lengths are no more than K. Alternatively, we could combine this with R^k_δ to obtain the relation $R^{k,K}_\delta$ where we require that $\sum_{i=0}^{k-1} d(x_i, x_{i+1}) \leq K$.

An immediate consequence of the definition of $\mathbf{L}^k_\delta$ is that for any metric space (X, d_X) and any threshold value $\delta \geq 0$, there exists $k \in \mathbb{N}$ such that $\mathbf{L}^k_\delta$ is equivalent to $\mathbf{SL}_\delta$ on (X, d_X). This may be summarized as saying that $\mathbf{SL}_\delta = \mathbf{L}^\infty_\delta$. Similarly, $\mathbf{ML}_\delta = \mathbf{L}^1_\delta$. The functor $\mathbf{L}^2_\delta$ is also known as "Cech clustering at scale δ" or sometimes "at 2δ".

Now let $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Cov}_\blacktriangleright$ be any (flat/non-hierarchical) clustering functor. We consider the two-point metric space Λ^1_ε with distance $\varepsilon \geq 0$ between the two points. Note that if $\varepsilon' \geq \varepsilon$ then there is a non-expansive mapping (morphism in $\mathbf{Met}$):

$$\Lambda^1_{\varepsilon'} \to \Lambda^1_\varepsilon.$$

Thus if $\mathbf{F}(\Lambda^1_\varepsilon)$ consists of two single point clusters then so does $\mathbf{F}(\Lambda^1_{\varepsilon'})$. On the other hand, if $\mathbf{F}(\Lambda^1_\varepsilon)$ is a single cluster, then so is $\mathbf{F}(\Lambda^1_{\varepsilon''})$ for any $\varepsilon'' \leq \varepsilon$.

We call $\mathbf{F}$ trivial if $\mathbf{F}(\Lambda^1_\varepsilon)$ is two single point clusters for all $\varepsilon \geq 0$ or if $\mathbf{F}(\Lambda^1_\varepsilon)$ is a single two point cluster for all $\varepsilon \geq 0$. One can easily show that in the former case $\mathbf{F}(X)$ is the cover by singletons for all (X, d_X) in $\mathbf{Met}$ and in the latter case $\mathbf{F}(X)$ is just the cover consisting of X itself for all (X, d_X). (Keep in mind that the cover $\mathbf{F}(X)$ is a flag cover.)

Thus if $\mathbf{F}$ is non-trivial there exists a number $\delta_\mathbf{F} \geq 0$ such that $\mathbf{F}(\Lambda^1_\varepsilon)$ is a single two point cluster if $\varepsilon < \delta_\mathbf{F}$ and two singleton clusters if $\varepsilon > \delta_\mathbf{F}$. The question of what happens when $\varepsilon = \delta_\mathbf{F}$ is a minor annoyance, and we will assume $\mathbf{F}(\Lambda^1_{\delta_\mathbf{F}})$ is a single two point cluster. The other case can be handled with some minor changes to our discussion.

Definition 7. Given a non-trivial clustering functor $\mathbf{F}$, we call $\delta_\mathbf{F}$ the *clustering parameter* for $\mathbf{F}$. □

Note that if $\mathbf{F}$ has clustering parameter $\delta_\mathbf{F}$ and (X, d_X) is any metric space with $x_1, x_2 \in X$ and $d_X(x_1, x_2) \leq \delta_\mathbf{F}$ then x_1 and x_2 lie in a common cluster (set of the cover) of $\mathbf{F}(X, d_X)$.

Theorem 8. *Suppose $\mathbf{F}$ is a non-trivial clustering functor $\mathbf{Met} \to \mathbf{Cov}_\blacktriangleright$ with clustering parameter $\delta_\mathbf{F}$. Then for any input space (X, d_X), the output of $\mathbf{F}$ refines the output of $\mathbf{SL}_{\delta_\mathbf{F}}$ and is refined by the output of $\mathbf{ML}_{\delta_\mathbf{F}}$.*

Proof. Suppose $x, y \in X$ such that $d_X(x, y) \leq \delta_\mathbf{F}$. Then there exists a morphism $\Lambda^1_{\delta_\mathbf{F}} \to (X, d_X)$ with image $\{x, y\}$. By the hypothesis, $\mathbf{F}$ merges $\Lambda^1_{\delta_\mathbf{F}}$ into one cluster, so there must be some cluster A in $\mathbf{F}(X, d_X)$ such that $x, y \in A$. Here we are using the functoriality of $\mathbf{F}$, which means we have a morphism

$$\mathbf{F}(\Lambda^1_{\delta_\mathbf{F}}) \to \mathbf{F}(X, d_X) = (X, \mathcal{C}_X)$$

in $\mathbf{Cov}_\blacktriangleright$, and our single two point cluster must refine the pullback of $\mathcal{C}_X$ to $\Lambda^1_{\delta_\mathbf{F}}$. Since $\mathbf{F}(X, d_X)$ is flag, if B is a maximal linkage component of (X, d_X) at scale $\delta_\mathbf{F}$ then B is contained in an element of $\mathbf{F}(X, d_X)$.

Now suppose x, y are elements of X such that x and y are in separate components of $\mathbf{SL}_{\delta_\mathbf{F}}(X, d_X)$. Then there exists a morphism $(X, d_X) \to \Lambda^1_\varepsilon$ for some $\varepsilon > \delta_\mathbf{F}$ which sends x and y to different points in Λ^1_ε. But this implies that x and y can never be in the same cluster in $\mathbf{F}(X, d_X)$. □

Note that this does not imply that the clusters in $\mathbf{F}(X, d_X)$ are unions of Rips clusters (i.e., $\mathbf{ML}_{\delta_\mathbf{F}}(X, d_X)$), which is false in general.

B. Functors on $\mathbf{Met}^{inj}$

Carlsson and Mémoli, after proving the uniqueness of $\mathbf{SL}$ as a functor $\mathbf{Met} \to \mathbf{Part}$, considered an expanded class of functors, those from $\mathbf{Met}^{\text{inj}} \to \mathbf{Part}$. In this section we consider some other clustering schemes in this context.

A number of overlapping clustering schemes have been suggested in the literature. Jardine and Sibson [21] proposed two "type B" methods that restricted the size of the overlap

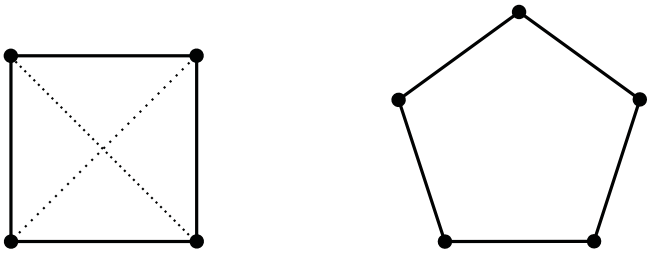

Fig. 2. Both these graphs are 2-connected, and if maximal would be clusters under $\mathbf{VL}^2$. However, the left graph would also be a cluster under $\mathbf{B}_k^c$, as the dotted edges are induced. Neither graph is a cluster under $\mathbf{B}_k$, since the maximal complete subgraphs are simply the edges, which overlap in sets of cardinality 1.

between clusters. We consider these two methods, along with two similar methods based on k-vertex and k-edge connectivity. The $\mathbf{B}_k$ clustering method is designed to prohibit overlaps of cardinality greater than or equal to k. One way to obtain it is by taking the maximally linked clusters for a given level δ, and repeatedly merging any two clusters that overlap in k or more points. Alternately, one may construct the threshold graph for a given δ, and then repeatedly add edges implied by the following property: if a and b are vertices, and there exists a complete subgraph S of size k such that both a and b are adjacent to every vertex in S, then a and b are adjacent. Then the $\mathbf{B}_k$ clusters are the maximal cliques of this graph. This requirement is relaxed in the coarser method $\mathbf{B}_k^c$, in which S need not be a complete subgraph, or even connected at all, i.e., a and b are each adjacent to a subset S of k points. Note that neither $\mathbf{B}_k$ nor $\mathbf{B}_k^c$ are functorial for $\mathbf{Met}$ (see figure 3 below).

We define the $\mathbf{VL}^k$ clustering methods as follows: Given a metric space (X, d_X) and $\delta \geq 0$, we construct the graph G with with vertices equal to the set X, where there is an edge between x and y if and only if $d_X(x, y) \leq \delta$. We call this graph the *δ-threshold graph* for (X, d_X). Then for any integer $k \geq 1$ construct the covering of X given by maximal k-vertex-connected subgraphs of G. We denote this clustering method $\mathbf{VL}^k$. Note that $\mathbf{SL}_\delta = \mathbf{VL}_\delta^1$. Further inspection shows that $\lim_{k\to\infty} \mathbf{VL}_\delta^k = \mathbf{ML}_\delta$. All three of these methods $\mathbf{B}_k, \mathbf{B}_k^c$ and $\mathbf{VL}^k$ are distinct, as Figure 2 shows.

The use of k-vertex connectivity in defining the $\mathbf{VL}^k$ clustering methods leads naturally to the idea of using k-edge-connectivity to separate clusters. Note that the maximal k-edge-connected subgraphs always form a partition of the vertices, unlike the maximal k-vertex-connected subgraphs. We will call this clustering method $\mathbf{EL}^k$. As with vertex connectivity, we also have that $\mathbf{EL}_\delta^1 = \mathbf{SL}_\delta$ and $\lim_{k\to\infty} \mathbf{EL}_\delta^k = \mathbf{ML}_\delta$ for all $\delta \geq 0$. In general, however, $\mathbf{EL}^k$ and $\mathbf{VL}^k$ will produce different results.

It is easy to see that each of these clustering methods fails to be functorial on $\mathbf{Met}$ for finite $k > 1$ and any $\delta \geq 0$. Consider the two spaces in Figure 3. For $\delta = 1$, X is grouped into a single cluster under the three overlap-restricting methods. However, the non-expansive mapping f takes X onto a metric space that has two clusters. The lack of functoriality stems from the restriction on numbers of overlapping points. Morphisms in $\mathbf{Met}$ may collapse several points into one, thus splitting a k-vertex-linked subgraph. Similarly, the two spaces in Figure 4 show that $\mathbf{EL}_\delta^k$ $(k > 1)$ is not functorial on $\mathbf{Met}$, with the problem again arising from the fact that multiple points can be collapsed into a single point. This motivates the consideration of the category $\mathbf{Met}^{\text{inj}}$ as in [10], which restricts morphisms to injective non-expansive maps.

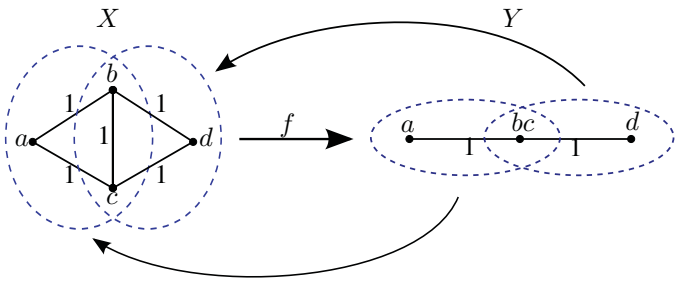

Fig. 3. For these metric spaces, all three overlapping clustering methods give the same result. The preimage of the $\mathbf{B}_2$, $\mathbf{B}_2^c$, or $\mathbf{VL}_1^2$ clusters in Y is finer than the clustering in X, so none of these methods is functorial over $\mathbf{Met}$.

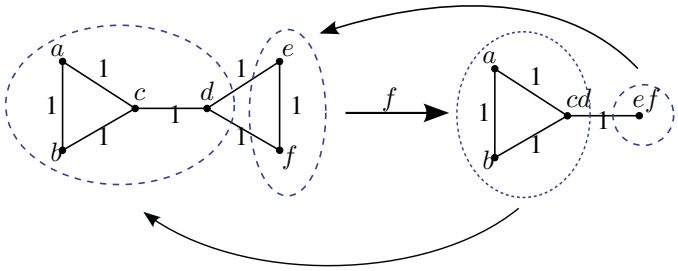

Fig. 4. The induced clusters from the non-expansive map f do not refine the original 2-edge-connected clusters, showing that $\mathbf{EL}_1^2$ is not functorial on $\mathbf{Met}$.

Theorem 9. *The mappings $\mathbf{B}_k, \mathbf{B}_k^c, \mathbf{VL}_\delta^k$ and $\mathbf{EL}_\delta^k$ from $\mathbf{Met}^{inj} \to \mathbf{Cov}_\blacktriangleright$ are functorial for each $k \geq 1$ and $\delta \geq 0$.*

Proof. Let $f\colon (X, d_X) \to (Y, d_Y)$ be a morphism in $\mathbf{Met}^{\text{inj}}$. If G_X and G_Y are the δ-threshold graphs associated to (X, d_X) and (Y, d_Y), then f induces a graph homomorphism $G_X \to G_Y$ since f is injective and non-expansive. Thus f preserves both edge and vertex connectedness, and the maximal edge or vertex-connected subsets of G_Y contain the images of the maximal connected subsets of G_X. In other words, $\mathbf{EL}_\delta^k(X, d_X)$ refines $f^{-1}\left(\mathbf{EL}_\delta^k(Y, d_Y)\right)$ and $\mathbf{VL}_\delta^k(X, d_X)$ refines $f^{-1}\left(\mathbf{VL}_\delta^k(X, d_X)\right)$, as desired.

The proof for $\mathbf{B}_k, \mathbf{B}_k^c$ follows directly from results of Jardine and Sibson [20], [21]. □

Theorem 10. *For every $\delta \geq 0$, there is a sequence of natural transformations*

$$\mathbf{ML}_\delta = \mathbf{VL}_\delta^\infty \to \cdots \to \mathbf{VL}_\delta^k \to \cdots \to \mathbf{VL}_\delta^1 = \mathbf{SL}_\delta$$

in the category of functors $\mathbf{Met}^{inj} \to \mathbf{Cov}_\blacktriangleright$.

Proof. Note that if $k \leq \ell$, the clustering given by $\mathbf{VL}^{\ell}_{\delta}$ always refines the clustering given by $\mathbf{VL}^{k}_{\delta}$. Then the identity maps from $\mathbf{ML}(X, d_X)$ to $\mathbf{SL}(X, d_X)$ are morphisms in $\mathbf{Cov}_{\blacktriangleright}$ for any X. □

The biconnected components of a graph can be computed in linear time; given a division of a graph into maximal k-connected subgraphs, these can be divided into $k+1$-connected subgraphs by finding all k-element vertex cuts. This can be done in polynomial time for each k, so the k-vertex connected components of a graph can be enumerated in polynomial time. (For more information see [24]). Constructing the adjacency graph for a given metric requires quadratic time in the number of points, so the $\mathbf{VL}^{k}_{\delta}$ clustering schemes can be calculated in polynomial time for any fixed k. However, the maximal clique problem is NP-complete, so no polynomial time algorithm is known to compute $\mathbf{ML}_{\delta}$ in general.

Note that the $\mathbf{EL}_{\delta}$ method is excisive in the sense of Carlsson–Mémoli [10] for each $\delta \geq 0$, so by Theorem 6.2 of *loc. cit.*, it is representable by a set of test metric spaces whose injections into X determine the clusters. However, it may be more efficiently calculated using one of several fast algorithms for finding maximal k-edge-connected subgraphs, such as those in [27], [11], and [1].

IV. Hierarchical Clustering

All of the parameterized flat clustering schemes we have considered generalize naturally to hierarchical clustering methods which we call *sieving functors* $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Sieve}$.

Theorem 11. *Suppose $\delta \leq \delta'$. Then for any k (including ∞) there is a natural transformation $\mathbf{L}^{k}_{\delta} \to \mathbf{L}^{k}_{\delta'}$.*

Proof. The theorem follows easily from the fact that for any X, the clustering given by $\mathbf{L}^{k}_{\delta}(X)$ refines that given by $\mathbf{L}^{k}_{\delta'}(X)$ whenever $\delta \leq \delta'$. □

Theorem 12. *Suppose $\mathbf{F}_t$ is a family of functors from $\mathbf{Met}$ to $\mathbf{Cov}_{\blacktriangleright}$ indexed by nonnegative real numbers t such that whenever $t \leq t'$, there is a natural transformation $\mathbf{F}_t \to \mathbf{F}_{t'}$. Then the map $\mathbf{F}\colon \mathbf{Met} \to \mathbf{Sieve}$ given by $(X, d_X) \mapsto (X, \theta_X)$, with $\theta(t) = \mathbf{F}_t(X, d_X)$ is a functor.*

The proof of Theorem 12 follows easily from the definition of a sieve, and we call a functor of this type a *sieving functor*. The two previous theorems then give us a family of hierarchical clustering schemes, $\{\mathbf{ML}, \ldots, \mathbf{L}^{k}, \ldots, \mathbf{SL}\}$. Note, however, that there are many more functorial hierarchical clustering schemes. A broader theoretical treatment is described in [13] where we work with sets having more general dissimilarity measures and provide a characterization of stable sieving functors.

V. Conclusion

We have discussed overlapping clustering schemes and the constraints that functoriality imposes. Functoriality, loosely speaking, enforces consistency in clustering across related data sets and typical operations on data sets such as sub-sampling and projection.

References

[1] Takuya Akiba, Yoichi Iwata, and Yuichi Yoshida. Linear-time enumeration of maximal k-edge-connected subgraphs in large networks by random contraction. In *Proceedings of the 22nd ACM international Conference on information and knowledge management*, pages 909–918, 2013.

[2] Hans-Jürgen Bandelt and Andreas W. M. Dress. Weak hierarchies associated with similarity measures: An additive clustering technique. *Bulletin of Mathematical Biology*, 51(1):133–166, 1989.

[3] Hans-Jürgen Bandelt and Andreas W. M. Dress. A canonical decomposition theory for metrics on a finite set. *Advances in Mathematics*, 92:47–105, 1992.

[4] Hans-Jürgen Bandelt and Andreas W. M. Dress. An order theoretic framework for overlapping clustering. *Discrete Mathematics*, 136(1–3):21–37, December 1994.

[5] Michael Barr and Charles Wells. *Category Theory for Computing Science*. Les Publications CRM, Montréal, 3rd edition, 1999.

[6] J.P. Barthélemy, F. Brucker, and C. Osswald. Combinatorial optimisation and hierarchical classifications. *Annals of Operations Research*, 153(1):179–214, September 2007.

[7] Patrice Bertrand. Set systems and dissimilarities. *European Journal of Combinatorics*, 21:727–743, 2000.

[8] Patrice Bertrand and Jean Diatta. Weak hierarchies: A central clustering structure. In Fuad Aleskerov, Boris Goldengorin, and Panos M. Pardalos, editors, *Clusters, Orders, and Trees: Methods and Applications*, Springer Optimization and Its Applications, pages 211–230. Springer New York, 2014.

[9] Gunnar Carlsson and Facundo Mémoli. Characterization, stability, and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.

[10] Gunnar Carlsson and Facundo Mémoli. Classifying clustering schemes. *Foundations of Computational Mathematics*, 13(1):221–252, 2013.

[11] Lijun Chang, Jeffrey Xu Yu, Lu Qin, Xuemin Lin, Chengfei Liu, and Weifa Liang. Efficiently computing k-edge connected components via graph decomposition. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, pages 205–216, 2013.

[12] Frédéric Chazal, Steve Oudot, Primoz Skraba, and Leonidas J. Guibas. Persistence-based clustering in Riemannian manifolds. *Journal of the Association of Computing Machinery*, 60(6), 2013.

[13] Jared Culbertson, Dan P. Guralnik, and Peter F. Stiller. Functorial clustering with overlaps. *Discrete Applied Mathematics*, 236:108–123, Feb 2018.

[14] Jean Diatta. One-to-one correspondence between indexed cluster structures and weakly indexed closed cluster structures. In Paula Brito, Patrice Bertrand, Guy Cucumel, and Francisco de Carvalho, editors, *Selected Contributions in Data Analysis and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 477–482. Springer Berlin Heidelberg, 2007.

[15] Andreas Dress, Vincent Moulton, Andreas Spillner, and Taoyang Wu. Obtaining splits from cut sets of tight spans. *Discrete Applied Mathematics*, 161:1409–1420, 2013.

[16] Andreas W. M. Dress, Katarina T. Huber, and Vincent Moulton. Totally split-decomposable metrics of combinatorial dimension two. *Annals of Combinatorics*, 5(1):99–112, 2001.

[17] Brian Everitt. *Cluster Analysis*. Wiley series in probability and statistics. Wiley, 5 edition, 2011.

[18] W Harvey, O Rübel, V Pascucci, P-T Bremer, and Y Wang. Enhanced topology-sensitive clustering by Reeb graph shattering. In Ronald Peikert, Helwig Hauser, Hamish Carr, and Raphael Fuchs, editors, *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications*, Mathematics and Visualization, pages 77–90. Springer Berlin Heidelberg, 2012.

[19] Melvin F. Janowitz. *Ordinal and Relational Clustering*, volume 10 of *Interdisciplinary Mathematical Sciences*. World Scientific, 2010.

[20] Nicholas Jardine and Robin Sibson. The construction of hierarchic and non-hierarchic classifications. *Computer Journal*, 11:117–184, 1968.

[21] Nicholas Jardine and Robin Sibson. *Mathematical Taxonomy*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. Wiley, 1971.

[22] Jon Kleinberg. An impossibility theorem for clustering. *Advances in Neural Information Processing Systems*, 15, 2002.

[23] Saunders Mac Lane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer, New York, 2nd edition, 1998.

[24] David W. Matula. k-blocks and ultrablocks in graphs. *Journal of Combinatorial Theory*, B(24):1–13, 1978.

[25] Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. Topological methods for the analysis of high dimensional data sets and 3D object recognition. In *Symposium on Point-Based Graphics*, pages 91–100, 2007.

[26] Bernd Sturmfels and Josephine Yu. Classification of six-point metrics. *The Electronic Journal of Combinatorics*, 11, 2004.

[27] Rui Zhou, Chengfei Liu, Jeffrey Xu Yu, Weifa Liang, Baichen Chen, and Jianxin Li. Finding maximal k-edge-connected subgraphs from a large graph. In *EDBT/ICDT Joint Conference*, March 2012.